Long sequence time-series forecasting (LSTF) plays a crucial role in this project's price prediction component. In essence it is the ability to capture precise long-range dependency coupling between output and input efficiently. Recent studies have shown increasing potential of using Transformers to increase forecasting performance. There are some key issues with this Transformers, namely quadratic time complexity and high memory usage.

Informer: An efficient transformer for LSTF

To address these limitations, Informer was introduced, it's an efficient transformer model tailored for LSTF. Unlike LSTMs which can struggle with longer sequences, Informer maintains high performance even as the prediction horizon increases.

One of the key innovations with Informer is the use of ProbSpare attention, which reduced the typical $O(N^2)$ time complexity of full attention to $O(N \log N)$. A significant improvement in training and inference speed without the sacrifice of accuracy.

Informer also incorporates a distillation mechanism within the encoder. This identifies and extracts the active data points while removing the passive ones between encoder layers. Leading to lower memory usage with a more compact model.

Informer additionally uses a generative inference strategy allowing the decoder to make predictions in parallel, a contrast to autoregressive decoding in traditional transformers. This is a further boost to the inference efficiency and overall efficiency of the method.

Unlike RNNs and LSTMs that process sequences sequentially, transformers operate with parallelised computation. Enabling them to capture complex non-linear dependencies across longer input sequences more effectively

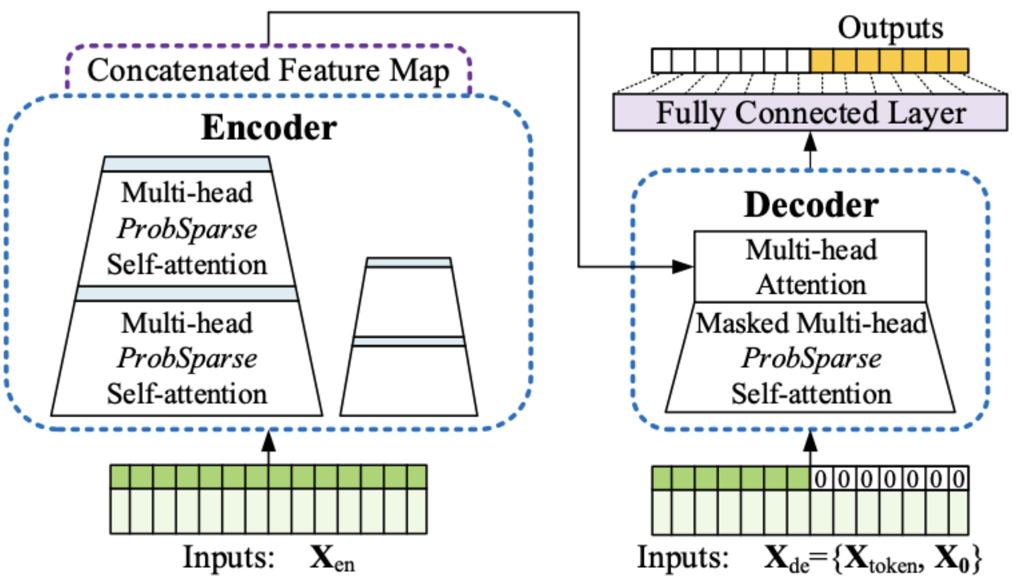

Figure 1: The Informer model overview. The blue space in the encoder is where the distilling takes place.

Methods		Informer		Informer†		LogTrans		Reformer		LSTMa		LSTnet	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh$_1$	24	**0.577**	**0.549**	0.620	0.577	0.686	0.604	0.991	0.754	0.650	0.624	1.293	0.901
	48	**0.685**	**0.625**	0.692	0.671	0.766	0.757	1.313	0.906	0.702	0.675	1.456	0.960
	168	**0.931**	**0.752**	0.947	0.797	1.002	0.846	1.824	1.138	1.212	0.867	1.997	1.214
	336	1.128	0.873	**1.094**	**0.813**	1.362	0.952	2.117	1.280	1.424	0.994	2.655	1.369
	720	**1.215**	**0.896**	1.241	0.917	1.397	1.291	2.415	1.520	1.960	1.322	2.143	1.380
ETTh$_2$	24	**0.720**	**0.665**	0.753	0.727	0.828	0.750	1.531	1.613	1.143	0.813	2.742	1.457
	48	**1.457**	**1.001**	1.461	1.077	1.806	1.034	1.871	1.735	1.671	1.221	3.567	1.687
	168	3.489	**1.515**	3.485	1.612	4.070	1.681	4.660	1.846	4.117	1.674	**3.242**	2.513
	336	2.723	1.340	2.626	**1.285**	3.875	1.763	4.028	1.688	3.434	1.549	**2.544**	2.591
	720	**3.467**	**1.473**	3.548	1.495	3.913	1.552	5.381	2.015	3.963	1.788	4.625	3.709
ETTm$_1$	24	0.323	**0.369**	**0.306**	0.371	0.419	0.412	0.724	0.607	0.621	0.629	1.968	1.170
	48	0.494	0.503	**0.465**	**0.470**	0.507	0.583	1.098	0.777	1.392	0.939	1.999	1.215
	96	**0.678**	0.614	0.681	**0.612**	0.768	0.792	1.433	0.945	1.339	0.913	2.762	1.542
	288	**1.056**	**0.786**	1.162	0.879	1.462	1.320	1.820	1.094	1.740	1.124	1.257	2.076
	672	**1.192**	**0.926**	1.231	1.103	1.669	1.461	2.187	1.232	2.736	1.555	1.917	2.941
Weather	24	**0.335**	**0.381**	0.349	0.397	0.435	0.477	0.655	0.583	0.546	0.570	0.615	0.545
	48	0.395	0.459	**0.386**	**0.433**	0.426	0.495	0.729	0.666	0.829	0.677	0.660	0.589
	168	**0.608**	**0.567**	0.613	0.582	0.727	0.671	1.318	0.855	1.038	0.835	0.748	0.647
	336	**0.702**	**0.620**	0.707	0.634	0.754	0.670	1.930	1.167	1.657	1.059	0.782	0.683
	720	**0.831**	**0.731**	0.834	0.741	0.885	0.773	2.726	1.575	1.536	1.109	0.851	0.757
ECL	48	0.344	**0.393**	**0.334**	0.399	0.355	0.418	1.404	0.999	0.486	0.572	0.369	0.445
	168	0.368	0.424	**0.353**	**0.420**	0.368	0.432	1.515	1.069	0.574	0.602	0.394	0.476
	336	0.381	**0.431**	0.381	0.439	**0.373**	0.439	1.601	1.104	0.886	0.795	0.419	0.477
	720	0.406	0.443	**0.391**	**0.438**	0.409	0.454	2.009	1.170	1.676	1.095	0.556	0.565
	960	**0.460**	**0.548**	0.492	0.550	0.477	0.589	2.141	1.387	1.591	1.128	0.605	0.599
Count		33		14		1		0		0		2	

Table 1: Multivariate long sequence time-series forecasting results on four datasets.

Informer shows a significant advantage in Mean Squared Error and Mean Absolute Error across the benchmark datasets. The performance margin outstrips more conventional models and Transformers.

In terms of computational efficiency, Informer outperforms other transformer-based models in both training and inference phases. The use of ProbSparse attention, encoder distillation and the parallel generative inference in decoding provide a significant runtime and memory improvement. All while maintain or improving the forecast accuracy.

Temporal Fusion Transformer

The Temporal Fusion Transformer (TFT) is another powerful model for multi-horizon and multivariate time series forecasting. It combines the strengths of recurrent neural networks (RNNs), attention mechanisms and interpretable design principles. Multi-horizon forecasting has many real-world applications, one of the key applications is in retail. In contrast to one-step-ahead predictions, multi-horizon forecasts provide access to estimates across the entire path at multiple steps in the future.

It would be useful in the context of price prediction modelling because practical multi-horizon applications commonly have access to a variety of data sources: information about the future, exogenous time series, static metadata and historical data. Some of the key mechanisms of TFT are static covariate encoders, gating mechanisms and temporal self-attention decoder. These facilitate interpretability, helping users identify globally important variables, persistent temporal patterns and significant events.

Static covariate encoders in TFT encode static features which allow the model to adjust the behaviour depending on the characteristics. Gating mechanisms control the flow of information in the architecture, allowing the model to dynamically select relevant inputs and skip irrelevant ones, leading to an improvement in generalisation and performance. So the model learns both long and short-term temporal relationships temporal processing is used with a sequence-to-sequence layer being employed for local processing. Whereas long-term dependencies are captured with an interpretable multi-head attention block.

TFT achieved state of the art forecasting performance in its tests. This is important because the tests involved grocery sales data from Kaggle. Forecasting log product sales 30 days into the future with 90 days of previous data as the base. This is relevant to our work and concepts can be taken over. TFT is also quite interpretable, in the attached paper three use cases were implemented. Examining the importance of each input variable in prediction, visualising temporal persistent patterns and identifying any regimes or events that lead to significant changes. (These can be found in more detail on page 20)

Transformers are an extremely powerful tool that can be utilised in the unique situation that our project is working on. As we can gain more data over a longer period this will become more prevalent. Transformers and particularly Informer which surpasses other transformers, not being plagued with issues and surpasses other modelling methods such as LSTM as it works significantly better with longer sequences as the prediction horizon increases. TFT has already been tested and proven itself in a similar situation to what DiscountMate's aims are and has higher interpretability.

References:

https://arxiv.org/abs/2012.07436
https://huggingface.co/blog/informer
https://huggingface.co/docs/transformers/en/model_doc/informer
https://medium.com/@bijit211987/transformers-like-informer-arrevolutionizing-time-series-forecasting-f4e4ebd7db1b
https://www.sciencedirect.com/science/article/abs/pii/S1566253523001355
https://www.youtube.com/watch?v=1xqvmr1gqJI
https://www.youtube.com/watch?v=aETHYkoJeNY
https://arxiv.org/abs/1912.09363
https://medium.com/dataness-ai/understanding-temporal-fusion-transformer-9a7a4fcde74b
https://medium.com/dataness-ai/forecasting-book-sales-with-temporal-fusion-transformer-dd482a7a257c

https://www.kaggle.com/code/tomwarrens/temporal-fusion-transformer-in-pytorch
https://pytorch-forecasting.readthedocs.io/en/latest/tutorials/stallion.html